# Bayes factors for linear regression

Rui Wang

Saturday 3rd November, 2018

## 1 Introduction

This note gives a review for Bayes factors for linear regression.

## 2 Mixture of $g$ prior

This section is adapted from Liang et al. (2008). Suppose $\mathbf{Y} \in \mathbb{R}^n$ is generated from the model

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \phi^{-1}\mathbf{I}_n)$.

Let $\mathbf{X}_\gamma \in \mathbb{R}^{n\times p_\gamma}$ be a submatrix of $\mathbf{X}$. Then the submodel $\mathcal{M}_\gamma$ is defined as

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \mathbf{X}_\gamma\boldsymbol{\beta}_\gamma + \boldsymbol{\varepsilon}.$$

The null model $\mathcal{M}_N$ is

$$\mathcal{M}_\gamma : \mathbf{Y} = \mathbf{1}_n \alpha + \boldsymbol{\varepsilon}.$$

We would like to compare $\mathcal{M}_\gamma$ with $\mathcal{M}_N$. Without loss of generality, we assume $\mathbf{1}_n^\top \mathbf{X}_\gamma = 0$. Under $\mathcal{M}_N$, the $g$ prior is

$$p(\alpha, \phi|\mathcal{M}_N) = \frac{1}{\phi}.$$

Under $\mathcal{M}_\gamma$, the $g$ prior is

$$\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top\mathbf{X}_\gamma)^{-1}), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

The joint pdf is

$$\begin{aligned}
&p(\mathbf{Y}, \alpha, \boldsymbol{\beta}_\gamma, \phi|\mathcal{M}_\gamma) = p(\mathbf{Y}|\alpha, \boldsymbol{\beta}_\gamma, \phi, \mathcal{M}_\gamma)p(\boldsymbol{\beta}_\gamma|\phi, \mathcal{M}_\gamma)p(\alpha|\phi, \mathcal{M}_\gamma)p(\phi|\mathcal{M}_\gamma)\\
=&(2\pi)^{-(n+p_\gamma)/2} g^{-p_\gamma/2} \phi^{(n+p_\gamma)/2-1} |\mathbf{X}_\gamma^\top \mathbf{X}_\gamma|^{1/2} \exp\left\{-\frac{n\phi}{2}(\bar{\mathbf{Y}} - \alpha)^2\right\}\\
&\exp\left\{-\frac{\phi(g+1)}{2g}\left\|\mathbf{X}_\gamma\left(\boldsymbol{\beta}_\gamma - \frac{g}{g+1}\hat{\boldsymbol{\beta}}_\gamma\right)\right\|^2 - \frac{\phi}{2(g+1)}\left\|\mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2 - \frac{\phi}{2}\left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\right\|^2\right\},
\end{aligned}$$

where $\bar{\mathbf{Y}} = n^{-1}\mathbf{1}_n^\top \mathbf{Y}$, $\hat{\boldsymbol{\beta}}_\gamma = (\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}\mathbf{X}_\gamma^\top \mathbf{Y}$.

Direct calculation yields

$$p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}},$$

where $R_\gamma^2 = 1 - \|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}} - \mathbf{X}_\gamma\hat{\boldsymbol{\beta}}_\gamma\|^2/\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\|^2$. Also, we have

$$p(\mathbf{Y}|\mathcal{M}_N) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2}\sqrt{n}} \left\|\mathbf{Y} - \mathbf{1}_n\bar{\mathbf{Y}}\right\|^{-(n-1)}.$$

Thus,

$$\mathrm{BF}[\mathcal{M}_\gamma : \mathcal{M}_N] = (1+g)^{(n-p_\gamma-1)/2}[1+g(1-R_\gamma^2)]^{-(n-1)/2}.$$

## 2.1 Choices of $g$

**Local empirical Bayes.** The local EB estimates a separate $g$ for each model $\mathcal{M}_\gamma$.

$$\hat{g}_\gamma^{\mathrm{EBL}} = \arg\max_{g\geq 0} p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}} = \max\{F_\gamma - 1, 0\},$$

where

$$F_\gamma = \frac{R_\gamma^2/p_\gamma}{(1-R_\gamma^2)/(n-1-p_\gamma)}$$

is the usual $F$ statistic for testing $\boldsymbol{\beta}_\gamma = 0$.

**Global empirical Bayes.** The global EB procedure assumes one common $g$ for all models.

$$\hat{g}_\gamma^{\mathrm{EBG}} = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma)p(\mathbf{Y}|\mathcal{M}_\gamma, g) = \arg\max_{g\geq 0} \sum_\gamma p(\mathcal{M}_\gamma)\frac{(1+g)^{(n-p_\gamma-1)/2}}{[1+g(1-R_\gamma^2)]^{(n-1)/2}}.$$

In general, this marginal likelihood is not tractable and does not provide a closed-form solution for $\hat{g}_\gamma^{\mathrm{EBG}}$. It can be computed by an EM algorithm, which is based on treating both the model indicator and the precision $\phi$ as latent data.

## 2.2 Mixtures of $g$ priors

Under $\mathcal{M}_\gamma$, the mixtures of $g$ prior take the form

$$\boldsymbol{\beta}_\gamma|g, \phi, \mathcal{M}_\gamma \sim \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1}), \quad \pi(g), \quad p(\alpha|\phi, \mathcal{M}_\gamma) \propto 1, \quad p(\phi|\mathcal{M}_\gamma) = \frac{1}{\phi}.$$

*Zellner-Siow Priors*

$$\pi(\boldsymbol{\beta}_\gamma|\phi) \propto \frac{\Gamma(p_\gamma)}{\pi^{p_\gamma/2}} \left|\frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi}\right|^{1/2} \left(1 + \boldsymbol{\beta}_\gamma^\top \frac{\mathbf{X}_\gamma^\top \mathbf{X}_\gamma}{n/\phi}\boldsymbol{\beta}_\gamma\right)^{-p_\gamma/2}$$

The Zellner-Siow priors can be represented as a mixture of $g$ priors with an Inv-Gamma(1/2,n/2) prior on $g$, namely,

$$\phi(\boldsymbol{\beta}_\gamma|\phi) \propto \int \mathcal{N}(0, \frac{g}{\phi}(\mathbf{X}_\gamma^\top \mathbf{X}_\gamma)^{-1})\pi(g)dg,$$

with

$$\pi(g) = \frac{(n/2)^{1/2}}{\Gamma(1/2)} g^{-3/2} e^{-n/(2g)}.$$

*Hyper-g priors*

$$\pi(g) = \frac{a-2}{2}(1+g)^{-a/2}\mathbf{1}_{(0,\infty)}(g), \quad a > 2.$$

Equivalently,

$$\frac{g}{1+g} \sim \text{Beta}(1, \frac{a}{2}-1).$$

The null-based Bayes factor is

$$\begin{aligned}\text{BF}[\mathcal{M}_\gamma : \mathcal{M}_N] =& \frac{a-2}{2}\int_0^\infty (1+g)^{(n-1-p_\gamma-a)/2}[1+(1-R_\gamma^2)g]^{-(n-1)/2}dg \\ =& \frac{a-2}{p_\gamma+a-2} \times {}_2F_1\left(\frac{n-1}{2}, 1; \frac{p_\gamma+a}{2}; R_\gamma^2\right),\end{aligned}$$

where ${}_2F_1(a, b; c; z)$ is the Gaussian hypergeometric function defined as

$${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}}{(1-tz)^a}dt.$$

*Beta prime prior* Maruyama and George (2011) proposed to use the beta prime prior for $g$:

$$\pi(g) = \frac{g^b(1+g)^{-a-b-2}}{B(a+1, b+1)}\mathbf{1}_{(0,\infty)}(g),$$

where $a > -1$, $b > -1$. Equivalently,

$$\frac{1}{1+g} \sim \text{Be}(a+1, b+1).$$

They observed that the Bayes factor has a closed form if we take

$$b = \frac{n-p_\gamma-5}{2} - a.$$

# 3 Intrinsic prior

Intrinsic prior, introduced by Berger and Pericchi (1996) and further developed by Moreno et al. (1998), is a method for objective Bayes hypothesis testing.

Suppose that two models are proposed for the data $\mathbf{x} = (x_1, \ldots, x_n)$. Under model $M_i, i = 1, 2$, the data are related to parameter $\theta_i$ by a density $f_i(\mathbf{x}|\theta_i)$. A noninformative prior for $\theta_i$ is denoted by $\pi_i^N(\theta_i)$, $i = 1, 2$. The conventional Bayes factor is defined as

$$B_{21}(\mathbf{x}) = \frac{m_2(\mathbf{x})}{m_1(\mathbf{x})} = \frac{\int_{\Theta_2} f_2(\mathbf{x}|\theta_2)\pi_2(\theta_2)d\theta_2}{\int_{\Theta_1} f_1(\mathbf{x}|\theta_1)\pi_1(\theta_1)d\theta_1}.$$

But the conventional Bayes factor suffers from arbitrary normalizing constant. To solve this problem, Berger and Pericchi (1996) proposed the intrinsic Bayes factor.

The intrinsic Bayes factor is based on training samples. This idea is to split the sample $\mathbf{x}$ into two parts as $\mathbf{x} = (x(l), x(n-l))$, where part $x(l)$, the training sample, is utilized to convert $\pi_i^N(\theta_i)$ into proper distributions,

$$\pi_i(\theta_i|x(l)) = \frac{f_i(x(l)|\theta_i)\pi_i^N(\theta_i)}{m_i^N(x(l))},$$

where $m_i^N(x(l)) = \int f_i(x(l)|\theta_i)\pi_i^N(\theta_i)d\theta_i$. With the remaining portion of the data $x(n-l)$, the Bayes factor is computed using the foregoing $\pi_i(\theta_i|x(l))$ as priors. The resulting partial Bayes factor is

$$B_{21}(x(n-l)|x(l)) := B_{21}(l) = B_{21}^N(\mathbf{x}) \cdot B_{12}^N(x(l)),$$

where

$$B_{12}^N(x(l)) = \frac{m_1^N(x(l))}{m_2^N(x(l))}.$$

Note that $B_{12}^N(l)$ does not depend on the arbitrary constants in $\pi_i^N(\theta_i)$. In addition, it is well defiend only if $x(l)$ is such that $0 < m_i^N(x(l)) < \infty$, $i = 1, 2$. If there is no subsample of $x(l)$ for which $0 < m_i^N(x(l)) < \infty$, $i = 1, 2$, then $x(l)$ is called a minimal training sample.

Berger and Pericchi (1996) suggested using a minimal training sample to compute $B_{21}(l)$ and to take an averate over all of the minumal training samples contained in the sample. This gives the arithmetic intrinsic Bayes factor (AIBF) of $M_2$ against $M_1$ as

$$B_{21}^{AI}(\mathbf{x}) = B_{21}^N(\mathbf{x}) \frac{1}{L} \sum_{i=1}^{L} B_{12}^N(x(l)),$$

where $L$ is the number of minimal training samples $x(l)$ contained in $\mathbf{x}$.

Other averaging methods can also be used. The geometric intrinsic Bayes factor (GIBF) is defined by

$$B_{21}^{GI}(\mathbf{x}) = B_{21}^N(\mathbf{x}) \left( \prod_{i=1}^{L} B_{12}^N(x(l)) \right)^{1/L}.$$

However, IBF is not an actual Bayes factor and is not coherent in many aspect. An important question about the AIBF is to know whether it corresponds to an actual Bayes factor for sensible priors. Such a prior, if it exists, is called an intrinsic prior. Berger and Pericchi (1996) define intrinsic priors by using an (asymptotic) imaginary training sample.

Let $\pi_1(\theta_1)$ and $\pi_2(\theta_2)$ be certain priors. The corresponding Bayes factor is

$$B_{21}(\mathbf{x}) = \frac{\int_{\Theta_2} f_2(\mathbf{x}|\theta_2)\pi_2(\theta_2)d\theta_2}{\int_{\Theta_1} f_1(\mathbf{x}|\theta_1)\pi_1(\theta_1)d\theta_1}.$$

The following approximation is valid in the standard situation.

$$B_{21} = B_{21}^N \cdot \frac{\pi_2(\hat{\theta}_2)\pi_1^N(\hat{\theta}_1)}{\pi_2^N(\hat{\theta}_2)\pi_1(\hat{\theta}_1)} \cdot (1 + o_P(1)),$$

where $\hat{\theta}_i$ are the MLE's under $M_i$, $i = 1, 2$. Equating $B_{21}$ and $B_{21}^{AI}(\mathbf{x})$, we have

$$\frac{\pi_2(\hat{\theta}_2)\pi_1^N(\hat{\theta}_1)}{\pi_2^N(\hat{\theta}_2)\pi_1(\hat{\theta}_1)} = \frac{1}{L}\sum_{i=1}^{L} B_{12}^N(x(l)) \quad \text{or} \quad \left(\prod_{i=1}^{L} B_{12}^N(x(l))\right)^{1/L}.$$

Suppose $M_1$ is the true model and $\theta_1$ is the true parameter. Letting $n \to \infty$, we have

$$\frac{\pi_2(\psi_2(\theta_1))\pi_1^N(\theta_1)}{\pi_2^N(\psi_2(\theta_1))\pi_1(\theta_1)} = \mathrm{E}_{\theta_1}^{M_1} B_{12}^N(x(l)) \quad \text{or} \quad \exp\left(\mathrm{E}_{\theta_1}^{M_1} \log B_{12}^N(x(l))\right), \tag{1}$$

where $\psi_2(\theta_1)$ is the limiting MLE of $\theta_2$. Similarly,

$$\frac{\pi_2(\theta_2)\pi_1^N(\psi_1(\theta_2))}{\pi_2^N(\theta_2)\pi_1(\psi_1(\theta_2))} = \mathrm{E}_{\theta_2}^{M_2} B_{12}^N(x(l)) \quad \text{or} \quad \exp\left(\mathrm{E}_{\theta_2}^{M_2} \log B_{12}^N(x(l))\right). \tag{2}$$

If $M_1$ is nested in $M_2$, then (1) is implicit in (2). A natural solution is given by

$$\pi_1^I(\theta_1) = \pi_1^N(\theta_1), \quad \pi_2^I(\theta_2) = \pi_2^N(\theta_2)\,\mathrm{E}_{\theta_2}^{M_2} B_{12}^N(x(l)).$$

Equivalently,

$$\pi_2^I(\theta_2|\theta) = \pi_2^N(\theta_2)\,\mathrm{E}_{\theta_2}^{M_2} \frac{f_1(x(l)|\theta_1)}{m_2^N(x(l))}.$$

### 3.1 Linear model

Casella and Moreno (2006) proposed a fully automatic Bayesian procedure for variable selection in normal regression models. The posterior probabilities are computed using intrinsic priors. Consider the standard normal regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^\top$ is the vector of observations, $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_k]$ is the $n \times k$ design matrix, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)^\top$ is the $k \times 1$ column vector of the regression coefficients, and $\boldsymbol{\varepsilon}$ is an error vector distributed as $\boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, \sigma^2\mathbf{I}_n)$. This is the full model for $\mathbf{y}$ and is denoted by $\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)$.

Let $\boldsymbol{\gamma}$ denote a vector of length $k$ with components equal to either 0 or 1, and let $\mathbf{Q}_{\boldsymbol{\gamma}}$ denote a $k \times k$ diagonal matrix with the elements of $\boldsymbol{\gamma}$ on the leading diagonal and 0 elsewhere. Because we want to include the intercept in every model, the first component of each $\boldsymbol{\gamma}$ is equal to 1. We let $\gamma$ denote the set of $2^{k-1}$ different configurations of $\boldsymbol{\gamma}$.

A submodel is written as $\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)$, where $\boldsymbol{\beta}_{\boldsymbol{\gamma}} = \mathbf{Q}_{\boldsymbol{\gamma}}\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$ is a configuration to be interpreted as $\gamma_i = 0$ if $\alpha_i = 0$ and 1 otherwise.

We have the Bayesian model

$$M_{\boldsymbol{\gamma}} : \{\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n), \pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})\}, \quad \boldsymbol{\gamma} \in \Gamma.$$

They used the full model method. The Bayes factor of a generic model $M_{\boldsymbol{\gamma}}$, when compared with the full model $M_{\mathbf{1}}$, is given by the ratio of marginal distributions

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y}, \mathbf{X}) = \frac{m_{\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{X})}{m_{\mathbf{1}}(\mathbf{y}, \mathbf{X})} = \frac{\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)\pi(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}}{\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma}.$$

Casella and Moreno (2006) considered the standard default prior on parameter $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$, giving the Bayesian model

$$M_{\boldsymbol{\gamma}} : \{\mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n), \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) = c_{\boldsymbol{\gamma}}/\sigma_{\boldsymbol{\gamma}}^2\}, \quad \boldsymbol{\gamma} \in \Gamma.$$

We first take an arbitrary but fixed point $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$ in the null space, and then find the intrinsic prior for $(\boldsymbol{\alpha}, \sigma)$ conditional on $(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})$. To do this, we note that a theoretical minimal training sample for this problem is a random vector $\mathbf{y}^{ts}$ of dimension $k+1$ such that it is $\mathcal{N}_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)$ distributed under the null model and is $\mathcal{N}_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)$ distributed under the full model. Here $\mathbf{Z}^{ts}$ represents a $(k+1)\times k$ unknown design matrix associated with $\mathbf{y}^{ts}$.

Therefore,

$$\pi^I(\boldsymbol{\alpha}, \sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}) = \pi^N(\boldsymbol{\alpha}, \sigma) \times \mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma}.$$

Note that

$$\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma = \int \left(\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)d\boldsymbol{\alpha}\right)\frac{c}{\sigma^2}d\sigma.$$

We have

$$\begin{aligned}
&\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)d\boldsymbol{\alpha} \\
=&\frac{1}{(2\pi)^{(k+1)/2}\sigma^{k+1}} \exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\cdot \\
&\int \exp\left\{-\frac{1}{2\sigma^2}\left\|\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\mathbf{y}^{ts} - \mathbf{Z}^{ts}\boldsymbol{\alpha}\right\|^2\right\}d\boldsymbol{\alpha} \\
=&\frac{1}{(2\pi)^{(k+1)/2}\sigma^{k+1}} \exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\cdot(2\pi)^{k/2}\sigma^k\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{-1/2} \\
=&\frac{1}{(2\pi)^{1/2}\sigma\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha}, \sigma)d\boldsymbol{\alpha}d\sigma \\
&=\int \frac{c}{(2\pi)^{1/2}\sigma^3\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}d\sigma \\
(\phi := \sigma^{-2}) &= \int \frac{c\phi^{3/2}}{(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}} \exp\left\{-\frac{\phi}{2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}\frac{1}{2}\phi^{-3/2}d\phi \\
&=\frac{c}{2(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}}\int \exp\left\{-\frac{\phi}{2}\left\|\left(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2\right\}d\phi \\
&=\frac{c}{(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\left\|(I - \mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top})\mathbf{y}^{ts}\right\|^2}.
\end{aligned}$$

With the above expression, we have

$$
\begin{aligned}
&\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma}\,\frac{N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)\pi^N(\boldsymbol{\alpha},\sigma)d\boldsymbol{\alpha}d\sigma}\\
=&c^{-1}(2\pi)^{1/2}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma}\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2 N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)\\
=&\frac{(2\pi)^{1/2}}{c}\left|\mathbf{Z}^{ts\top}\mathbf{Z}^{ts}\right|^{1/2}\int\left\|\left(I-\mathbf{Z}^{ts}(\mathbf{Z}^{ts\top}\mathbf{Z}^{ts})^{-1}\mathbf{Z}^{ts\top}\right)\mathbf{y}^{ts}\right\|^2 N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma^2_{\boldsymbol{\gamma}}\mathbf{I}_n)N_{k+1}(\mathbf{y}^{ts}|\mathbf{Z}^{ts}\boldsymbol{\alpha},\sigma^2\mathbf{I}_n)d\mathbf{y}^{ts}.
\end{aligned}
$$

To compute this integral, we use the following lemma.

**Lemma 1.**

$$
\int_{\mathbb{R}^n}\left(\mathbf{y}^\top\mathbf{K}\mathbf{y}\prod_{i=1}^{2}\mathcal{N}_n(\mathbf{y}|\mathbf{X}\theta_i,\sigma_i^2\mathbf{I}_n)\right)d\mathbf{y}=\frac{\sigma_2^2\,\mathrm{tr}(\mathbf{K})|\mathbf{X}^\top\mathbf{X}|^{-1/2}}{(2\pi\sigma_1^2)^{(n-k)/2}(1+\sigma_2^2/\sigma_1^2)^{(n-k+2)/2}}\mathcal{N}_k(\theta_2|\theta_1,(\sigma_1^2+\sigma_2^2)(\mathbf{X}^\top\mathbf{X})^{-1}),
$$

*where* $\mathbf{K}$ *is an* $n\times n$ *symmetric matrix,* $\mathbf{X}$ *is an* $n\times k$ *matrix of rank* $k$ *such that* $\mathbf{K}\mathbf{X}=0$*.*

*Proof.*

$$
\begin{aligned}
&\int_{\mathbb{R}^n} \left( \mathbf{y}^\top \mathbf{K}\mathbf{y} \prod_{i=1}^{2} \mathcal{N}_n(\mathbf{y}|\mathbf{X}\theta_i, \sigma_i^2 \mathbf{I}_n) \right) d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K}\mathbf{y} \exp\left\{ -\frac{1}{2\sigma_1^2}\|\mathbf{y}-\mathbf{X}\theta_1\|^2 - \frac{1}{2\sigma_2^2}\|\mathbf{y}-\mathbf{X}\theta_2\|^2 \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K}\mathbf{y} \exp\left\{ -\frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2} \left( \|\mathbf{y}\|^2 - 2\mathbf{y}^\top \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right) \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K}\mathbf{y} \exp\left\{ -\frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2} \left( \left\| \mathbf{y} - \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 - \left\| \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right) \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 + \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} \left\| \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right\} \cdot \\
&\int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K}\mathbf{y} \exp\left\{ -\frac{1}{2}\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2} \left\| \mathbf{y} - \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right\} d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 + \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} \left\| \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right\} \cdot \\
&(2\pi)^{n/2} \left( \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \right)^{n/2} \int_{\mathbb{R}^n} \mathbf{y}^\top \mathbf{K}\mathbf{y} \mathcal{N}_n \left( \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right), \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}\mathbf{I}_n \right) d\mathbf{y} \\
=&\frac{1}{(2\pi)^n \sigma_1^n \sigma_2^n} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 + \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} \left\| \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right\} \cdot \\
&(2\pi)^{n/2} \left( \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2} \right)^{n/2+1} \operatorname{tr}(\mathbf{K}) \\
=&\frac{\sigma_1^2\sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2}(\sigma_1^2+\sigma_2^2)^{n/2+1}} \exp\left\{ -\frac{1}{2\sigma_1^2}\theta_1^\top \mathbf{X}^\top \mathbf{X}\theta_1 - \frac{1}{2\sigma_2^2}\theta_2^\top \mathbf{X}^\top \mathbf{X}\theta_2 + \frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2} \left\| \mathbf{X} \left( \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}\theta_1 + \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}\theta_2 \right) \right\|^2 \right\} \\
=&\frac{\sigma_1^2\sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2}(\sigma_1^2+\sigma_2^2)^{n/2+1}} \exp\left\{ -\frac{1}{2(\sigma_1^2+\sigma_2^2)}(\theta_1-\theta_2)^\top \mathbf{X}^\top \mathbf{X}(\theta_1-\theta_2) \right\} \\
=&\frac{\sigma_1^2\sigma_2^2 \operatorname{tr}(\mathbf{K})}{(2\pi)^{n/2}(\sigma_1^2+\sigma_2^2)^{n/2+1}} \cdot (2\pi)^{k/2}(\sigma_1^2+\sigma_2^2)^{k/2}|\mathbf{X}^\top \mathbf{X}|^{-1/2} \mathcal{N}_n \left( \theta_2|\theta_1, (\sigma_1^2+\sigma_2^2)(\mathbf{X}^\top \mathbf{X})^{-1} \right) \\
=&\frac{\sigma_1^2\sigma_2^2 \operatorname{tr}(\mathbf{K})|\mathbf{X}^\top \mathbf{X}|^{-1/2}}{(2\pi)^{(n-k)/2}(\sigma_1^2+\sigma_2^2)^{(n-k)/2+1}} \mathcal{N}_n \left( \theta_2|\theta_1, (\sigma_1^2+\sigma_2^2)(\mathbf{X}^\top \mathbf{X})^{-1} \right).
\end{aligned}
$$

□

Using the above Lemma, we have

$$\begin{aligned}
&\int \left\| \left( I - \mathbf{Z}^{ts} (\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1} \mathbf{Z}^{ts\top} \right) \mathbf{y}^{ts} \right\|^2 N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) d\mathbf{y}^{ts} \\
=&\frac{\sigma^2 |\mathbf{Z}^{ts\top} \mathbf{Z}^{ts}|^{-1/2}}{(2\pi\sigma_{\boldsymbol{\gamma}}^2)^{1/2} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).
\end{aligned}$$

It follows that

$$\begin{aligned}
&\mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^N(\boldsymbol{\alpha}, \sigma) d\boldsymbol{\alpha} d\sigma} \\
=&\frac{(2\pi)^{1/2}}{c} \left| \mathbf{Z}^{ts\top} \mathbf{Z}^{ts} \right|^{1/2} \int \left\| \left( I - \mathbf{Z}^{ts} (\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1} \mathbf{Z}^{ts\top} \right) \mathbf{y}^{ts} \right\|^2 N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) d\mathbf{y}^{ts} \\
=&\frac{\sigma^2}{c \sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}) \\
=&\pi^N(\boldsymbol{\alpha}, \sigma) \times \mathrm{E}_{\mathbf{y}^{ts}|\boldsymbol{\alpha},\sigma} \frac{N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n)}{\int N_{k+1}(\mathbf{y}^{ts} | \mathbf{Z}^{ts} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^N(\boldsymbol{\alpha}, \sigma) d\boldsymbol{\alpha} d\sigma} \\
=&\frac{1}{\sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2)(\mathbf{Z}^{ts\top} \mathbf{Z}^{ts})^{-1}).
\end{aligned}$$

**Proposition 1.** *The conditional intrinsic prior is*

$$\pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}) = \frac{1}{\sigma_{\boldsymbol{\gamma}} (1 + \sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}} \mathcal{N}_k(\boldsymbol{\alpha} | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2 + \sigma_{\boldsymbol{\gamma}}^2) \mathbf{W}^{-1}),$$

*where* $\mathbf{W} = \mathbf{Z}^{ts\top} \mathbf{Z}^{ts}$.

A way of assesing $\mathbf{W}^{-1}$ is to use the original idea of the arithmetic intrinsic Bayes factor. This entails averaging over all possible training samples of minimal size contained in the sample. This would give the matrix

$$\mathbf{W}^{-1} = \frac{1}{L} \sum_{l=1}^{L} (\mathbf{Z}^\top(l) \mathbf{Z}(l))^{-1},$$

where $\{\mathbf{Z}(l), l = 1, \ldots, L\}$ is the set of all submatrices of $\mathbf{X}$ of order $(k+1) \times k$ of rank $k$.

For the data $(\mathbf{y}, \mathbf{X})$, the Bayes factor for comparing models $M_{\boldsymbol{\gamma}}$ and $M_{\mathbf{1}}$ with the intrinsic priors $\{\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\beta}}), \pi^I(\boldsymbol{\alpha}, \sigma)\}$ has the formal expression

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y}, \mathbf{X}) = \frac{\int \mathcal{N}_n(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2 \mathbf{I}_n) \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) d\boldsymbol{\beta}_{\boldsymbol{\gamma}} d\sigma_{\boldsymbol{\gamma}}}{\int \mathcal{N}_n(\mathbf{y} | \mathbf{X} \boldsymbol{\alpha}, \sigma^2 \mathbf{I}_n) \pi^I(\boldsymbol{\alpha}, \sigma | \boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) \pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}) d\boldsymbol{\alpha} d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}} d\sigma_{\boldsymbol{\gamma}}}.$$

In what follows we partition the design matrix $\mathbf{X}$ as $\mathbf{X} = (\mathbf{X}_{0\boldsymbol{\gamma}} | \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})$, where $\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}$ contains the column $j$ of $\mathbf{X}$ if the configuration $\boldsymbol{\gamma}$ is such that $\gamma_j = 1$. Therefore, the dimension of $\mathbf{X}_{1\boldsymbol{\gamma}}$ is $n \times k_{\boldsymbol{\gamma}}$, where $k_{\boldsymbol{\gamma}} = \sum_{i=1}^{k} \gamma_i$.

**Proposition 2.** *The Bayes factor is given by*

$$B_{\boldsymbol{\gamma}\mathbf{1}}(\mathbf{y},\mathbf{X}) = \left(|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{1/2} (\mathbf{y}^\top (\mathbf{I}_n - \mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y})^{(n-k_{\boldsymbol{\gamma}}+1)/2} I_{\boldsymbol{\gamma}}\right)^{-1},$$

*where* $\mathbf{H}_{\boldsymbol{\gamma}} = \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top$,

$$\begin{aligned}
I_{\boldsymbol{\gamma}} =& \int_0^{\pi/2} \frac{d\varphi}{|\mathbf{A}_{\boldsymbol{\gamma}}(\varphi)|^{1/2} |\mathbf{B}(\varphi)|^{1/2} E_{\boldsymbol{\gamma}}(\varphi)^{(n-k_{\boldsymbol{\gamma}}+1)/2}},\\
\mathbf{B}(\varphi) =& (\sin^2\varphi)\mathbf{I}_n + \mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top,\\
\mathbf{A}_{\boldsymbol{\gamma}}(\varphi) =& \mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{B}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}},\\
E_{\boldsymbol{\gamma}}(\varphi) =& \mathbf{y}^\top \left(\mathbf{B}^{-1}(\varphi) - \mathbf{B}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\mathbf{A}_{\boldsymbol{\gamma}}^{-1}(\varphi)\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top \mathbf{B}^{-1}(\varphi)\right)\mathbf{y}.
\end{aligned}$$

*Proof.* For the numerator, we have

$$\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}}^2\mathbf{I}_n)\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\boldsymbol{\beta}_{1\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-\hat{\boldsymbol{\beta}}_{\mathbf{1}\boldsymbol{\gamma}})\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{1}{(2\pi)^{n/2}\sigma_{\boldsymbol{\gamma}}^n}(2\pi)^{k_{\boldsymbol{\gamma}}/2}\sigma_{\boldsymbol{\gamma}}^{k_{\boldsymbol{\gamma}}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\exp\left\{-\frac{1}{2\sigma_{\boldsymbol{\gamma}}^2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\sigma_{\boldsymbol{\gamma}}\\
=&\int \frac{\phi^{(n-k_{\boldsymbol{\gamma}})/2}}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\exp\left\{-\frac{\phi}{2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}c_{\boldsymbol{\gamma}}\phi\left(\frac{1}{2\phi^{3/2}}\right)d\phi\\
=&\int \frac{c_{\boldsymbol{\gamma}}}{2}\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\phi^{(n-k_{\boldsymbol{\gamma}}+1)/2-1}\exp\left\{-\frac{\phi}{2}\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2\right\}d\phi\\
=&\frac{c_{\boldsymbol{\gamma}}}{2}\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)\left(\frac{2}{\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^2}\right)^{(n-k_{\boldsymbol{\gamma}}+1)/2}\\
=&\frac{c_{\boldsymbol{\gamma}}}{\sqrt{2}}\frac{1}{(\pi)^{(n-k_{\boldsymbol{\gamma}})/2}}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2}\Gamma((n-k_{\boldsymbol{\gamma}}+1)/2)\|(\mathbf{I}_n-\mathbf{H}_{\boldsymbol{\gamma}})\mathbf{y}\|^{-(n-k_{\boldsymbol{\gamma}}+1)}.
\end{aligned}$$

Now we deal with the denominator

$$\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha},\sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}}.$$

We have

$$\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha},\sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}},\sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma\\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\mathcal{N}_k(\boldsymbol{\alpha}|\boldsymbol{\beta}_{\boldsymbol{\gamma}},(\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{W}^{-1})\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}d\boldsymbol{\alpha}d\sigma\\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}},(\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top+\sigma^2\mathbf{I}_n)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}d\sigma.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\alpha}, \sigma^2\mathbf{I}_n)\pi^I(\boldsymbol{\alpha}, \sigma|\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\boldsymbol{\alpha}d\sigma d\boldsymbol{\beta}_{\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}} \\
=&\int \mathcal{N}_n(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\gamma}}, (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}\pi^N(\boldsymbol{\beta}_{\boldsymbol{\gamma}}, \sigma_{\boldsymbol{\gamma}})d\sigma d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}d\sigma_{\boldsymbol{\gamma}} \\
=&\int \left(\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}\right)\frac{1}{\sigma_{\boldsymbol{\gamma}}(1+\sigma^2/\sigma_{\boldsymbol{\gamma}}^2)^{3/2}}\frac{c_{\boldsymbol{\gamma}}}{\sigma_{\boldsymbol{\gamma}}^2}d\sigma d\sigma_{\boldsymbol{\gamma}}.
\end{aligned}$$

Let $\tilde{\boldsymbol{\Sigma}} = (\sigma^2+\sigma_{\boldsymbol{\gamma}}^2)\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sigma^2\mathbf{I}_n$. Then

$$\begin{aligned}
&\int \mathcal{N}_n\left(\mathbf{y}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}, \tilde{\boldsymbol{\Sigma}}\right)d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}} \\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\int \exp\left\{-\frac{1}{2}(\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}})^\top\tilde{\boldsymbol{\Sigma}}^{-1}(\mathbf{y}-\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}})\right\}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}} \\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}\cdot \\
&\int \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{y}\right)^\top\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}\left(\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}}-(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{y}\right)\right\}d\boldsymbol{\beta}_{\mathbf{1}\boldsymbol{\gamma}} \\
=&\frac{1}{(2\pi)^{n/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}\cdot \\
&(2\pi)^{k_{\boldsymbol{\gamma}}/2}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{-1/2} \\
=&\frac{1}{(2\pi)^{(n-k_{\boldsymbol{\gamma}})/2}|\tilde{\boldsymbol{\Sigma}}|^{1/2}|\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}|^{1/2}}\exp\left\{-\frac{1}{2}\mathbf{y}^\top\left(\tilde{\boldsymbol{\Sigma}}^{-1}-\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}(\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}})^{-1}\mathbf{X}_{\mathbf{1}\boldsymbol{\gamma}}^\top\tilde{\boldsymbol{\Sigma}}^{-1}\right)\mathbf{y}\right\}.
\end{aligned}$$

Let $\sigma_{\boldsymbol{\gamma}} = \rho\cos\varphi$, $\sigma = \rho\sin\varphi$, where $\rho \in (0, +\infty)$, $\varphi \in (0, \pi/2)$. Then

$$\tilde{\boldsymbol{\Sigma}} = \rho^2(\mathbf{X}\mathbf{W}^{-1}\mathbf{X}^\top + \sin^2\varphi\mathbf{I}_n) = \rho^2\mathbf{B}(\varphi),$$

□

# 4 Normal-inverse-gamma (NIG) prior

Zhou and Guan (2018)

Consider the testing problem in linear regression with independent normal errors:

$$\begin{aligned}
H_0 &: \mathbf{Y}|\mathbf{a}, \tau \sim \mathcal{N}(\mathbf{W}\mathbf{a}, \tau^{-1}\mathbf{I}_n), \\
H_1 &: \mathbf{Y}|\mathbf{a}, \mathbf{b}, \tau \sim \mathcal{N}(\mathbf{W}\mathbf{a}+\mathbf{L}\mathbf{b}, \tau^{-1}\mathbf{I}_n),
\end{aligned}$$

where $\mathbf{W}$ is a full-rank $n \times q$ matrix representing the nuisance covariates, including a column of $\mathbf{1}_n$. $\mathbf{L}$ is an $n \times p$ matrix representing the covariates of interest.

NIG prior:

$$\begin{aligned}\mathbf{a}|\tau &\sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_a),\\ \mathbf{b}|\tau &\sim \mathcal{N}(0, \tau^{-1}\mathbf{V}_b),\\ \tau &\sim \text{Gamma}(\kappa_1/2, 2/\kappa_2).\end{aligned}$$

Here

$$\pi(\tau) = \frac{(\kappa_2/2)^{\kappa_1/2}}{\Gamma(\kappa_1/2)}\tau^{\kappa_1/2-1}\exp\left\{-\frac{\kappa_2\tau}{2}\right\}$$

Then

$$\begin{aligned}&f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)\\ =&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+p+q+\kappa_1)/2-1}}{(2\pi)^{(n+p+q)/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)}\exp\left\{-\frac{\tau}{2}\left(\|\mathbf{Y}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2+\mathbf{a}^\top\mathbf{V}_a^{-1}\mathbf{a}+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}+\kappa_2\right)\right\}.\end{aligned}$$

Let $\mathbf{H_W} = \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. Then

$$\begin{aligned}\|\mathbf{Y}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2 =&\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2+\|\hat{\mathbf{Y}}-\mathbf{W}\mathbf{a}-\mathbf{L}\mathbf{b}\|^2\\ =&\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2+\|\mathbf{H_W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})-\mathbf{W}\mathbf{a}\|^2+\|(\mathbf{I}_n-\mathbf{H_W})(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2.\end{aligned}$$

We have

$$\begin{aligned}&\int_{\mathbb{R}^q}\exp\left\{-\frac{\tau}{2}\left(\|\mathbf{H_W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})-\mathbf{W}\mathbf{a}\|^2+\mathbf{a}^\top\mathbf{V}_a^{-1}\mathbf{a}\right)\right\}d\mathbf{a}\\ =&\int_{\mathbb{R}^q}\exp\Big\{-\frac{\tau}{2}\Big(\Big\|(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{1/2}\Big[\mathbf{a}-(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})\Big]\Big\|^2\\ &-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\mathbf{H_W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2\Big)\Big\}d\mathbf{a}\\ =&(2\pi)^{q/2}\tau^{-q/2}|\mathbf{W}\mathbf{W}^\top+\mathbf{V}_a^{-1}|^{-1/2}\\ &\exp\left\{-\frac{\tau}{2}\left(-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\mathbf{H_W}(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})\|^2\right)\right\}.\end{aligned}$$

Thus,

$$\begin{aligned}&\int_{\mathbb{R}^q}f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}\\ =&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+p+\kappa_1)/2-1}}{(2\pi)^{(n+p)/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top+\mathbf{V}_a^{-1}|^{1/2}}\\ &\exp\Big\{-\frac{\tau}{2}\Big(\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b}\|^2\\ &+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}+\kappa_2\Big)\Big\}.\end{aligned}$$

Note that

$$\begin{aligned}&\int_{\mathbb{R}^p}\exp\left\{-\frac{\tau}{2}\left(-(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})^\top\mathbf{W}(\mathbf{W}^\top\mathbf{W}+\mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top(\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b})+\|\hat{\mathbf{Y}}-\mathbf{L}\mathbf{b}\|^2+\mathbf{b}^\top\mathbf{V}_b^{-1}\mathbf{b}\right)\right\}d\mathbf{b}\\ &=(2\pi)^{p/2}\tau^{-p/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L}+\mathbf{V}_b^{-1}|^{-1/2}\exp\left\{-\frac{\tau}{2}\hat{\mathbf{Y}}^\top\left(\mathbf{C}-\mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L}+\mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}}\right\},\end{aligned}$$

where $\mathbf{C} = \mathbf{I}_n - \mathbf{W}(\mathbf{W}^\top \mathbf{W} + \mathbf{V}_a^{-1})^{-1}\mathbf{W}^\top$. Thus,

$$\begin{aligned}
&\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}d\mathbf{b}\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\tau^{(n+\kappa_1)/2-1}}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\\
&\exp\Big\{-\frac{\tau}{2}\Big(\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2\Big)\Big\}.
\end{aligned}$$

Thus,

$$\begin{aligned}
&\int_0^\infty\int_{\mathbb{R}^p}\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},\mathbf{b},\tau)\pi(\mathbf{a}|\tau)\pi(\mathbf{b}|\tau)\pi(\tau)d\mathbf{a}d\mathbf{b}d\tau\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\Gamma((n+\kappa_1)/2)}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}|\mathbf{V}_b|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\\
&\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2}{2}\right)^{-(n+\kappa_1)/2}.
\end{aligned}$$

Under $H_0$, that is $\mathbf{V}_b \to 0$, we have

$$\begin{aligned}
&\int_0^\infty\int_{\mathbb{R}^q} f(\mathbf{Y}|\mathbf{a},0,\tau)\pi(\mathbf{a}|\tau)\pi(\tau)d\mathbf{a}d\tau\\
=&\frac{(\kappa_2/2)^{\kappa_1/2}\Gamma((n+\kappa_1)/2)}{(2\pi)^{n/2}|\mathbf{V}_a|^{1/2}\Gamma(\kappa_1/2)|\mathbf{W}\mathbf{W}^\top + \mathbf{V}_a^{-1}|^{1/2}}\\
&\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}_0\|^2 + \hat{\mathbf{Y}}_0^\top\mathbf{C}\hat{\mathbf{Y}}_0 + \kappa_2}{2}\right)^{-(n+\kappa_1)/2}.
\end{aligned}$$

Thus,

$$\mathrm{BF} = \frac{1}{|\mathbf{V}_b|^{1/2}|\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1}|^{1/2}}\left(\frac{\|\mathbf{Y}-\hat{\mathbf{Y}}\|^2 + \hat{\mathbf{Y}}^\top\left(\mathbf{C} - \mathbf{C}\mathbf{L}(\mathbf{L}^\top\mathbf{C}\mathbf{L} + \mathbf{V}_b^{-1})^{-1}\mathbf{L}^\top\mathbf{C}\right)\hat{\mathbf{Y}} + \kappa_2}{\|\mathbf{Y}-\hat{\mathbf{Y}}_0\|^2 + \hat{\mathbf{Y}}_0^\top\mathbf{C}\hat{\mathbf{Y}}_0 + \kappa_2}\right)^{-(n+\kappa_1)/2}$$

# References


Berger, J. O. and Pericchi, L. R. (1996). The intrinsic bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91(433):109–122.

Casella, G. and Moreno, E. (2006). Objective bayesian variable selection. *Journal of the American Statistical Association*, 101(473):157–167.

Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

Maruyama, Y. and George, E. I. (2011). Fully bayes factors with a generalized g -prior. *Ann. Statist.*, 39(5):2740–2765.

Moreno, E., Bertolino, F., and Racugno, W. (1998). An intrinsic limiting procedure for model selection and hypotheses testing. *Journal of the American Statistical Association*, 93(444):1451–1460.

Zhou, Q. and Guan, Y. (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association*, 0(0):1–10.